

Optimal design and analysis of clinical trials with repeated measures

Citation for published version (APA):

Winkens, B. (2005). *Optimal design and analysis of clinical trials with repeated measures*. [Doctoral Thesis, Maastricht University]. <https://doi.org/10.26481/dis.20051123bw>

Document status and date:

Published: 01/01/2005

DOI:

[10.26481/dis.20051123bw](https://doi.org/10.26481/dis.20051123bw)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

The purpose of the present thesis is to offer guidelines to practitioners for the design and analysis of randomized clinical trials with repeated measures, assuming either a linearly divergent or second-order polynomial treatment effect. The sample size, the number of repeated measures per subject and the allocation of time-points at which the measures are taken can be chosen optimally at the design stage. A design that yields maximally efficient treatment effect estimators is then called an optimal design. Optimal design issues are discussed in Chapters 1, 4, 5 and 6, while the selection of an analysis method is discussed in Chapters 1, 2 and 3.

Chapter 1 is introductory. It discusses two treatment effect patterns: linearly divergent and second-order polynomial treatment effects. The former is characterized by a linearly increasing difference between the treatment groups over time. Alternatively, if the difference first increases relatively fast and then stabilizes or even decreases over time, the treatment effect may be represented by a second-order polynomial function. The chapter proceeds with the description of a linear mixed effects model. This model is very useful for the analysis of repeated measures data, because it is very flexible in modelling different components of variation, like random effects, serial correlation and measurement error. Different covariance structures may be obtained with these components. The generalized least squares (GLS) method is the most commonly used estimation method for the fixed effects in this model. Chapter 1 ends by describing simple analysis methods, like analysis of covariance (ANCOVA), and introducing the log-transformation to normalize positively skewed data. The difference between a linear mixed effects and an ANCOVA model is explained. Both models assume that the data are normally distributed. This assumption is, however, often violated for positively skewed data, e.g. costs of medical care. A log-transformation may then be used to normalize the data.

If each subject is measured twice, i.e. once before treatment (pre-treatment) and once after treatment (post-treatment), the pre- and post-treatment measures may be treated as a bivariate outcome variable in a linear mixed effects model or the pre-treatment may be treated as a covariate in an analysis of covariance (ANCOVA) model. In Chapter 2, ANCOVA models are compared with linear mixed effects models with respect to the estimation of a overall treatment effect, i.e. difference in post-treatment expectations between the groups. Four different situations are considered:

- I. the regression slopes and residual post-treatment variances in an ANCOVA model are both homogeneous across groups,
- II. heterogeneous regression slopes and homogeneous residual post-treatment variances,
- III. homogeneous regression slopes and heterogeneous residual post-treatment variances,
- IV. heterogeneous regression slopes and residual post-treatment variances.

If the slopes are homogeneous (situation I and III), ANCOVA is preferred to the linear mixed effects approach, because it yields unbiased treatment effect and variance estimators. If the slopes are heterogeneous and the residual variances are homogeneous (situation II), a linear mixed effects model makes assumptions that are too complicated to apply with the present software, and ANCOVA should be used. Finally, if the slopes and residual variances are both heterogeneous (situation IV), a linear mixed effects model is preferred to an ANCOVA model, since it takes into account the variability of the pre-treatment mean estimator.

Many statistical methods of analysis assume that the outcome variable is normally distributed. This assumption is violated for positively skewed data, which are often encountered in practice. For example, costs for medical care and medical recovery times can only take on positive values and are often affected by positive outliers (high costs or slow recovery). A log-transformation may then be used to normalize the data. Chapter 3 discusses the effect of log-transforming positively skewed data on the power of a two-sample *t*-test, analysis of variance (ANOVA) or covariance (ANCOVA). It is shown that the required sample size to detect a treatment effect with a certain power may be decreased by log-transforming the data. This decrease in sample size can be substantial if the coefficient of variation (CV), i.e. the standard deviation divided by the mean, of the untransformed data is large.

Chapter 4 discusses the effect of adding intermediate measures on the variance of the generalized least squares (GLS) treatment effect estimator for six different covariance structures, assuming a linearly divergent treatment effect, equidistant time-points and either a fixed sample size or a fixed study budget. It shows that this effect strongly depends on the covariance structure of the repeated measures. The decrease in variance by adding intermediate, equally spaced measures is relatively large for a fixed sample size if

- a) the covariance structure is compound symmetric or
- b) the structure is close to compound symmetry and the correlation between two repeated measures does not exceed 0.80 or
- c) the correlation does not exceed 0.60, if the time-lag between the measures goes to zero.

Summary

For a fixed budget and a linear cost function, the design with two repeated measures per subject, i.e. one at the start and one at the end of the study, often yield highly efficient (small variance) treatment effect estimators. Only in the special case where the covariance structure resembles compound symmetry and the cost per subject is ten or more times larger than the cost per repeated measure, are more than two repeated measures required to obtain this goal. The covariance structure is, however, often unknown at the design stage of a study. In that case, an optimal design based on a first-order auto-regressive structure with measurement error (AR1+ME) should be used to be robust against misspecification of the covariance structure.

Repeated measures are often planned to be equidistant, but this does not always yield highly efficient treatment effect estimators. In Chapter 5, the optimal allocation of time-points is calculated and compared with equidistant time-points for a small number of repeated measures, different covariance structures and linearly divergent treatment effects. The effect of adding optimally spaced measures (at the expense of subjects) is also investigated. The assumed covariance structure is crucial for the results. Firstly, for covariance structures resembling compound symmetry (the correlation between repeated measures decreases relatively slowly with increasing time-lag), each subject should be measured about five times and equidistant time-points do not yield highly efficient treatment estimators. The first measure should then be taken at the start and the other measures should be obtained at the end of the study, close to each other in time. Secondly, if the covariance structure resembles first-order auto-regression (the correlation between repeated measures decreases relatively fast with increasing time-lag), two repeated measures, i.e. one at the start and one at the end of the study, are sufficient to obtain highly efficient treatment effect estimators. Finally, if the covariance structure is unknown at the design stage of a study, an optimal design based on a first-order auto-regressive structure with measurement error is preferable in terms of robustness against misspecification of the covariance structure. In this case, each subject should be measured three times, i.e. once at the start and twice at the end of the study.

Chapter 6 deals with the optimal selection of sample size and the number of repeated measures per subject for clinical trials with equidistant time-points and second-order polynomial treatment effects. Two optimality criteria are considered:

- 1) a D_s -optimality criterion, i.e. minimize the area of the simultaneous confidence region of the linear and quadratic treatment effect estimators, and
- 2) a c -optimality criterion, i.e. minimize the variance of a linear combination of the linear and quadratic treatment effect estimator. A linear combination is chosen so that it

corresponds to the difference between the areas under the curves of the treatment and control groups (*AUC*).

The effect of adding intermediate measures (at the expense of subjects) on the efficiency of the treatment effect estimators depends strongly on the assumed covariance structure and is hardly affected by the choice between the two optimality criteria. Given a fixed number of subjects, the increase in efficiency by adding intermediate, equally spaced measures is large for covariance structures close to compound symmetry and small for structures that resemble first-order auto-regression. For a fixed study budget and a linear cost function, the design with only three equally spaced measures is often highly efficient. Only if the structure resembles compound symmetry and the cost per subject is eight or more times larger than the cost per repeated measure, each subject should be measured more than three times. On the other hand, if the covariance structure is unknown, the design with three equally spaced measures should be used. Given a design with three repeated measures and a second-order polynomial treatment effect, equidistant time-points are D_s -optimal and highly efficient in terms of the *AUC*-optimality criterion.

Finally, Chapter 7 (epilogue) summarizes the results of this thesis and offers guidelines regarding the design and method of analysis for randomized clinical trials with linearly divergent or second-order polynomial treatment effects.

Samenvatting in het Nederlands (Dutch summary)

Als resultaat van dit proefschrift worden in hoofdstuk 7 (epiloog) een aantal praktische richtlijnen aangeboden m.b.t. het opzetten en analyseren van gerandomiseerde klinische onderzoeken met herhaalde metingen. We nemen hierbij aan dat het effect van een behandeling of medicijn linear toeneemt in de tijd of te beschrijven is met een kwadratische functie. Het aantal individuen, het aantal metingen per individu en de tijdstippen waarop de metingen genomen worden kunnen optimaal gekozen worden tijdens de opzetfase van een onderzoek. Een onderzoeksopzet is optimaal wanneer deze leidt tot maximaal efficiënte behandelingsschatters. Hoofdstukken 1, 4, 5 en 6 gaan over het optimaal opzetten van gerandomiseerde klinische onderzoeken, terwijl de selectie van een analyse-methode wordt besproken in de hoofdstukken 1, 2 en 3.

Hoofdstuk 1 is inleidend van aard. Hierin worden twee behandelingseffecten behandeld: lineair divergerende en kwadratische. Het eerste behandelingseffect wordt gekenmerkt door een verschil tussen twee behandelingsgroepen dat lineair toeneemt in de tijd. Anderzijds, als het verschil eerst relatief snel toeneemt en daarna stabiliseert of zelfs afneemt, dan spreken we van een kwadratisch behandelingseffect. In het hoofdstuk wordt vervolgens een lineair mixed effects model beschreven. Het analyseren van herhaalde metingen data is erg gemakkelijk met dit model, omdat het erg flexibel is in het modelleren van verschillende variatiecomponenten, zoals random effecten, seriële correlatie en meetfouten. Verschillende covariantie structuren kunnen verkregen worden met deze variatiecomponenten en de generaliseerde kleinste kwadraten (Engels: generalized least squares; GLS) schattingsmethode is de meest gebruikte methode om de fixed effecten mee te schatten in een lineair mixed effects model. Hoofdstuk 1 eindigt met het beschrijven van eenvoudige analyse methoden, zoals covariantie-analyse (Engels: analysis of covariance; ANCOVA), en met het introduceren van de log-transformatie om positief scheve data te normaliseren. Het verschil tussen een lineair mixed effects en een ANCOVA model wordt verklaard. Beide modellen nemen aan dat de data normaal verdeeld zijn. Deze aanname is echter vaak geschonden voor positief scheve data, b.v. de kosten voor medische zorg. Een log-transformatie kan dan gebruikt worden om de data te normaliseren.

Als elk individu twee keer is gemeten, d.w.z. een keer voor behandeling (voormeting) en een keer na behandeling (nameting), dan kan de voor- en nameting als een bivariate uitkomstvariabele opgenomen worden in een lineair mixed effects model of de voormeting kan als covariaat meegenomen worden in een covariantie-analyse (ANCOVA) model. In

hoofdstuk 2 worden ANCOVA modellen vergeleken met lineair mixed effects modellen t.a.v. het schatten van een algeheel behandelingseffect, d.w.z. het verwachte verschil tussen de groepen op de nameting. Vier verschillende situaties worden behandeld:

- I. de regressiehellingen en residuele varianties in een ANCOVA model zijn beiden homogeen over de groepen,
- II. heterogene regressiehellingen en homogene residuele varianties,
- III. homogene regressiehellingen en heterogene residuele varianties,
- IV. heterogene regressiehellingen en residuele varianties.

Als de hellingen homogeen zijn (situaties I en III), dan verdient ANCOVA de voorkeur boven de lineair mixed effects methode, omdat ANCOVA leidt tot zuivere behandelingseffect- en variantieschatters. Als de hellingen heterogeen zijn en de residuele varianties homogeen (situatie II), dan maakt een lineair mixed effects model aannames die te ingewikkeld zijn om gemodelleerd te worden met de huidige software. ANCOVA zal dan gebruikt moeten worden. Tenslotte, als de hellingen en residuele varianties beiden heterogeen zijn over de groepen (situatie IV), dan verdient een lineair mixed effects model de voorkeur boven een ANCOVA model, want alleen de eerste houdt rekening met de variabiliteit van de schatter van het populatiegemiddelde op de voormeting.

Veel statistische analysemethoden nemen aan dat de uitkomstvariabele normaal verdeeld is. Deze aanname is geschonden voor positief scheve data. Kosten voor medische zorg en medisch hersteltijden zijn voorbeelden van positief scheve data, omdat deze alleen positieve waarden kunnen aannemen en vaak uitschieters hebben naar boven (hoge kosten of langzaam herstel). Een log-transformatie kan dan gebruikt worden om deze data te normaliseren. Hoofdstuk 3 bespreekt het effect van het log-transformeren van positief scheve data op het vermogen van een twee-steekproeven *t*-toets, variantie-analyse (Engels: analysis of variance; ANOVA) of covariantie-analyse (ANCOVA). In dit hoofdstuk is aangetoond dat het vereiste aantal proefpersonen om een behandelingseffect te detecteren met een bepaald vermogen kan dalen door een log-transformatie. Deze daling in steekproefgrootte kan erg substantieel zijn als de variatie-coëfficiënt (Engels: coefficient of variation; CV) van de ongetransformeerde data groot is, d.w.z. de standaard deviatie is groot t.o.v. het gemiddelde.

Hoofdstuk 4 bespreekt het effect van het toevoegen van tussenliggende metingen op de variantie van de generaliseerde kleinste kwadraten (GLS) schatter van het behandelingseffect, waarbij zes covariantie structuren worden bekeken. Er wordt aangenomen dat het behandelingseffect lineair divergerend is, de tijdstippen waarop gemeten wordt op gelijke afstand van elkaar liggen en dat de steekproefgrootte of het studie-budget vastligt. We

hebben aangetoond dat dit effect erg afhangt van de covariantie-structuur van de herhaalde metingen. De daling in variantie door het toevoegen van tussenliggende, op gelijke afstand genomen metingen is relatief groot bij een vooropgestelde steekproefgrootte, als

- a) de covariantie structuur compound symmetrisch is of
- b) de structuur dichtbij een compound symmetrische structuur ligt en de correlatie tussen twee herhaalde metingen altijd kleiner is dan 0.80 of
- c) de correlatie nooit groter is dan 0.60, zelfs niet als de tijdsspanne tussen twee metingen naar 0 gaat.

Als het budget vastligt en de kostenfunctie lineair is, dan leidt het onderzoeksopzet waarbij elk individu twee keer wordt gemeten, d.w.z. een keer aan het begin en een keer aan het eind van het onderzoek, tot erg efficiënte schatters van het behandelingseffect. Alleen in het speciale geval waarin de covariantie-structuur lijkt op compound symmetrie en de kosten per individu tien of meer keer hoger is dan de kosten per meting, zijn meer dan twee metingen nodig om dit doel te bereiken. Echter de covariantie-structuur is vaak onbekend in de opzetfase van een onderzoek. In dat geval moet een optimaal onderzoeksopzet, dat gebaseerd is op een eerste-orde auto-regressieve covariantie-structuur met meetfout, gebruikt worden om bestand/robuust te zijn tegen het verkeerd specificeren van de covariantie-structuur.

Herhaalde metingen worden vaak gepland met gelijke afstand tussen de metingen, maar dit leidt niet altijd tot heel efficiënte schatters van het behandelingseffect. In hoofdstuk 5 is de optimale ligging van tijdstippen uitgerekend en vergeleken met gelijk-verdeelde tijdstippen (gelijke afstand tussen de metingen), waarbij verschillende covariantie-structuren worden bekeken. Verder is aangenomen dat het aantal metingen per persoon beperkt is en het behandelingseffect lineair divergerend is. Het effect van het toevoegen van optimaal gelegen tussenmetingen (ten koste van het aantal proefpersonen) is ook onderzocht. De aangenomen covariantie-structuur bleek cruciaal te zijn voor de resultaten. Ten eerste, als de covariantie-structuur dichtbij compound symmetrie ligt, d.w.z. de correlatie tussen herhaalde metingen daalt relatief langzaam met het toenemen van de tijdsspanne tussen de metingen, dan moet elk individu ongeveer vijf keer gemeten worden, waarbij de eerste meting aan het begin genomen moet worden en de andere vier, kort na elkaar, aan het eind van het onderzoek. Ten tweede, als de covariantie-structuur dichtbij eerste-orde auto-regressie ligt, d.w.z. de correlatie tussen herhaalde metingen daalt relatief snel met het toenemen van de tijdsspanne tussen de metingen, dan zijn twee metingen per individu, d.w.z. een meting aan het begin en een aan het eind van het onderzoek, voldoende om heel efficiënte schatters van het behandelingseffect te krijgen. Tenslotte, als de covariantie-structuur onbekend is in de opzetfase van een onderzoek,

dan verdient een optimale onderzoeksopzet gebaseerd op een eerste-order auto-regressieve covariantie-structuur met meetfout de voorkeur t.a.v. de robuustheid tegen het verkeerd specificeren van de covariantie-structuur. In dit geval moet elk individu drie keer gemeten worden, een keer aan het begin en twee keer, kort na elkaar, aan het eind van het onderzoek.

Hoofdstuk 6 behandelt de optimale selectie van de steekproefgrootte en het aantal metingen per individu voor klinische experimenten, waarin de herhaalde metingen met gelijke afstand ertussen genomen zijn en het behandelingseffect een kwadratische functie van de tijd is. Twee optimaliteitscriteria zijn gebruikt:

- 1) een D_s -optimaliteitscriterium, d.w.z. minimaliseer de oppervlakte van het simultane betrouwbaarheidsgebied van de lineaire en kwadratische behandelingseffectschatters,
- 2) een c -optimaliteitscriterium, d.w.z. minimaliseer de variantie van een lineaire combinatie van de lineaire en kwadratische behandelingseffectschatters. Een lineaire combinatie is gekozen zodat deze overeenstemt met het verschil tussen de oppervlakten onder de curves (Engels: areas under the curves; AUC) van de behandelings- en controlegroep.

Het effect van het toevoegen van tussengelegen metingen (ten koste van het aantal proefpersonen) op de efficiëntie van de schatters van het behandelingseffect hangt sterk af van de aangenomen covariantie-structuur en wordt nauwelijks beïnvloed door de keuze tussen de twee optimaliteitscriteria. Gegeven een vast aantal proefpersonen, de toename in efficiëntie door het toevoegen van tussengelegen, op gelijke afstand van elkaar genomen metingen is groot wanneer de covariantie-structuur dichtbij compound symmetrie ligt en deze is klein wanneer de structuur dichtbij eerste-order auto-regressie ligt. Als het studiebudget vastligt en de kostenfunctie lineair is, dan is de onderzoeksopzet, waarbij elk individu slechts drie keer gemeten is, vaak erg efficiënt. Alleen wanneer de structuur dichtbij compound symmetrie ligt en de kosten per individu acht of meer keer hoger is dan de kosten per meting, dan moet elk individu vaker dan drie keer gemeten worden. Anderzijds, als de covariantie-structuur onbekend is, dan moet de onderzoeksopzet met drie herhaalde metingen en gelijke afstand tussen de metingen gebruikt worden. Gegeven een onderzoeksopzet met drie herhaalde metingen en een kwadratisch behandelingseffect, gelijk-verdeelde tijdstippen zijn D_s -optimaal en heel efficiënt in termen van de AUC-optimaliteitscriterium.

Tenslotte, hoofdstuk 7 (epiloog) vat de resultaten van dit proefschrift samen en biedt guidelines aan die betrekking hebben op het opzetten en analyseren van gerandomiseerde klinische experimenten waarin het behandelingseffect lineair divergerend of een kwadratische functie van de tijd is.